

R3 P25 ①

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

H04N 7/52

H04J 3/06

[12] 发明专利申请公开说明书

[21] 申请号 99805668.5

[43] 公开日 2001 年 7 月 4 日

[11] 公开号 CN 1302513A

[22] 申请日 1999.12.21 [21] 申请号 99805668.5

[30] 优先权

[32] 1999.1.6 [33] EP [31] 99200027.3

[86] 国际申请 PCT/EP99/10306 1999.12.21

[87] 国际公布 WO00/41400 英 2000.7.13

[85] 进入国家阶段日期 2000.10.31

[71] 申请人 皇家飞利浦电子有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 R·陶里

W·R·T·滕卡特

[74] 专利代理机构 中国专利代理(香港)有限公司

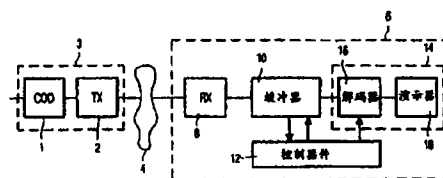
代理人 梁永 陈景峻

权利要求书 2 页 说明书 14 页 附图页数 8 页

[54] 发明名称 用于发送多媒体信号的传输系统

[57] 摘要

在一个通讯系统中,多媒体信号在编码器(1)中编码,并通过包 开关网络(4)传送到终端(6)。终端(6)包括一接收器(8),接收器的输出连接到接收缓冲器(210)。接收缓冲器(210)的输出施加到包含解码器(216)和演示器(218)的演示装置(214)。为了处理包开关网络(4)的延迟变化,建议根据多媒体信号的传输延迟改变多媒体信号的演示速度。这通过控制器(212)完成,它确定缓冲器(210)中包的数量并相应地调节多媒体信号的解码速率和回放速率。



ISSN 1008-4274

权 利 要 求 书

1. 用于复制多媒体信号并包含有用于向用户演示多媒体信号的演示设备的装置，其特征在于其中的装置平台包括用于确定表征携带有多媒体信号的包的到达延迟的延迟测量结果的延迟确定设备，并且其中的演示设备根据上述延迟测量结果改变演示速度。

2. 权利要求1所述的装置，其特征在于其中的多媒体信号包括音频信号，而且其中的演示设备可改变音频信号的演示速度，并且基本上不改变音频信号的可觉察语调。

3. 权利要求2所述的装置，其特征在于其中的音频信号由多个包括至少由其振幅和频率描述的多个信号的节表征，并且其中的演示设备可根据上述延迟测量结果改变上述节的延续时间。

4. 权利要求1所述的装置，其特征在于其中的演示设备包括具有用于确定表征延迟测量结果和参考值之差的差分信号的比较设备的控制设备，而且其中的演示设备包括用于根据差分值调节演示速度的调节设备。

5. 权利要求4所述的装置，其特征在于其中的演示设备包括用于根据差分值的变化而适应性调节参考值的适应设备。

6. 权利要求1所述的装置，其特征在于其中的多媒体信号包括视频信号。

7. 权利要求6所述的装置，其特征在于其中的视频信号由至少一个对象表征，而且其中的演示设备通过调节视频信号中至少一个对象的移动速度来改变演示速度。

8. 权利要求1所述的装置，其特征在于其中的多媒体信号包括至少两个分量，其中的延迟测量结果表征上述至少两个分量之间的时序差，而且其中的演示设备可改变演示速度以减小上述时序差。

9. 用于复制多媒体信号的方法，上述方法包括向用户演示多媒体信号，其特征在于该方法进一步包括确定表征携带有多媒体信号的包的到达延迟的延迟测量结果，而且该方法包括根据上述延迟测量结果改变演示速度。

10. 权利要求9所述的方法，其特征在于其中的多媒体信号包括音频信号，而且该方法包括改变音频信号的演示速度，并且基本上不改变音频信号的可觉察语调。

11. 权利要求 10 所述的方法，其特征在于其中的音频信号由多个包括多个至少由其振幅和频率描述的波形的节表征，而且该方法包括根据上述延迟测量结果改变上述节的延续时间。

12. 权利要求 9 所述的方法，其特征在于其中的多媒体信号包括
5 视频信号。

13. 权利要求 12 所述的方法，其特征在于其中的视频信号由至少一个对象表征，而且该方法包括通过调节视频信号中至少一个对象的移动速度来改变演示速度。

说明书

用于发送多媒体信号的传输系统

5 本发明涉及到一套用于复制多媒体信号、包括向用户演示多媒体信号的演示设备的装置。本发明也涉及到一种复制多媒体信号的方法。

从 V. Hardman 等人于 1995 年 5 月 4 日发表在 URL 为 <http://www.isoc.org/HMP/PAPER/2070/html/paper.html> 的 ISOC 网站中的文章“通过互联网使用的可靠音频”中可以知道这样一套系统。

10 在上述文章中描述的系统都用于通过诸如，举例来说，互联网，ATM 网络或 MPEG-2 传输流的包交换网传输多媒体信号，如音频和视频信息。

15 通过包交换网实时传输多媒体信号涉及到的主要问题是会发生包丢失，包延迟和包延迟扩散。包丢失可使用用于将包的不完整序列在演示给用户之前完整化的重构技术来解决。

包延迟扩散的解决方法是使用大的接收缓冲器以便能够一直有可演示给用户的包。为使之成为可能，接收缓冲器应足够大，以便可以应付可能发生的最大延迟扩散。由此得到多媒体信号在演示给用户之前有相当大的延迟。

20 多媒体信号的大的延迟在完全双工通信系统，如网络电话系统和类似于电视会议系统和网络游戏的多方系统中尤其是个问题。

本发明的目的是提供一套根据前述的传输系统，其中总的端到端延迟被基本消除了。

25 为实现上述目的，依照本发明的传输系统的特征在于其中的第二平台包括用于确定携带有多媒体信号的包的到达延迟的延迟确定设备，而且其中的演示设备能够根据携带有多媒体信号的包的上述到达延迟改变演示速度。

30 通过确定包延迟并根据上述包延迟确定演示速度，在第二平台中可以使用较小尺寸的缓冲器来解决延迟扩散问题。由于第二平台中的缓冲器尺寸较小，因而可以基本消除总的端到端延迟。

实验显示，大约 240% 的演示速度变化都几乎不被用户注意到。

可以看到，在 H. Sanneck 等人提供给 IEEE Globecom 219296

Conference, London, November 218-222, 219296 并发表在 Global Internet '296Conference Record, pp.248-252 的文章“一种新的音频包丢失掩盖技术”中, 提供了一种通过对原始信号的时间拉伸来重构丢失的包的方法。不过可以看到, 上述文章并没有提及将时间拉伸作为一种工具用来消除用于传输多媒体信号的通信系统中的端到端延迟。

可以看到, 本发明构想不仅仅适用于可将抖动引入多媒体信号中的多媒体信号通过网络的传输, 而且也适用于其中多媒体信号显示有抖动的所有情形。

这个情况的第一个例子是需要可在编程处理器中计算多媒体信号中的内容。计算时间取决于多媒体的实际内容, 所以在精确的时间间隔之后不一定都有可利用的多媒体信号。这是, 举例来说, 在运行多任务操作系统的计算机中和在现在所有的计算机游戏中常常遇到的多媒体信号计算涉及渲染精美的 3D 图像时的情形。第二个例子是从存储器件, 比如 CD-ROM 或硬盘中提取多媒体信号。

访问时间随读出头的实际位置的不同而发生变化, 因此会将抖动引入多媒体信号中。

如果演示速度取决于多媒体信号的可利用性, 就可更平稳地演示多媒体信号。

本发明的一个具体实施例的特征在于多媒体信号中包括一个音频信号, 其中的演示设备用于改变音频信号的演示速度, 而基本不会改变音频信号的可觉察语调。

改变音频信号的演示速度且不改变其语调, 会降低改变了演示速度的可听度。从以前的技术中可以知道改变音频信号演示速度且不改变音频信号语调的一些方法。在上述 Globecom 的文章中给出了这样一个例子。

依据本发明的通信系统的一个推荐具体实施例的特征在于其中的音频信号由多个包括多个至少用其振幅和频率描述的音频信号的节表征, 而且其中的演示设备用于根据包的上述可利用性来改变上述节的持续时间。

音频信号的这种表征方法的使用使得很容易地改变演示速度, 并且不改变音频信号的语调。在这种表征方法中, 音频信号的基频是由

表征该信号的信号的属性决定的，重构音频信号时用到的节的长度决定了演示速度。

当重构装置中用到的节的长度大于节的标称长度时，回放演示速度要低于原始的演示速度。

- 5 当重构装置中用到的节的长度小于节的标称长度时，回放演示速度要高于原始的演示速度。

10 本发明的一个更进一步的实施例的特征在于其中的演示设备包括具有用于确定表征延迟测量结果和参考值之间差值的差分信号的比较设备，而且其中的演示设备包括用于根据该差值调节演示速度的调节设备。

这个具体实施例提供了一种由延迟测量结果确定演示速度的容易而有效的方法。

本发明的一个更进一步的实施例的特征在于其中的演示设备包括用于根据差值变化适应性调节参考值的适应设备。

- 15 通过根据差值的变化改变参考值，可以根据多媒体信号中出现的抖动的实际总量确定平均缓冲器尺寸。如果抖动高，参考值就会有一个高的值，因此在缓冲器中就会有大量的包。如果抖动低，参考值就会有一个低值，因此在缓冲器中就会有少量的包。

- 20 在此方法中，缓冲器中实际尺寸永远不会大于应付多媒体信号中出现的抖动的实际总量所需的尺寸。

本发明的一个更进一步的实施例可用于多媒体信号包括视频信号的情况，其特征在于该视频信号由至少一个对象表征，而且其中的演示设备可通过调节视频信号中至少一个对象的运动速度来改变演示速度。

- 25 本发明的这个具体实施例可用于就像 MPEG-4 视频信号那样由数个分立对象表征的视频信号。在此类视频信号中，通过调节一个或多个对象的运动速度可以轻易地改变演示速度。这种改变演示速度的方法几乎不会被该装置的用户察觉到。

- 30 本发明的一个更进一步的实施例的特征在于其中的多媒体信号包括至少两个分量，其中的延迟测量结果表征了上述至少两个分量之间的时序差，而且其中的演示设备用于改变演示速度，以消除上述时序差。

本发明也适用于同步一个多媒体信号的两个或多个分量。延迟测量结果表征了两个分量之间的时序差。该时序差可以，举例来说，从与每个分量一起包含在多媒体信号中的时间标记中推出。

现在结合附图解释本发明。

5 图 1 是依据本发明的一个通信系统的方框图。

图 2 是如图 1 所示的通信系统的控制器 212。

图 3 是如图 1 所示的通信系统中使用的控制器 212 的一个替代具体实施例。

图 4 是如图 1 所示的通信系统中使用的编码器 1 的方框图。

10 图 5 是如图 1 所示的通信系统中使用的解码器 216 的方框图。

图 6 更详细地显示出解码器 216 中使用的谐波语音合成器 294。

图 7 是合成帧长是常数时，谐波语音合成器 294 中的不同波形。

图 8 是当两个相邻合成帧之间的合成帧长变化时，谐波语音合成器 294 中的不同波形。

15 图 9 更详细地显示出解码器 216 中使用的无声语音合成器 296。

图 10 是如图 1 所示的系统中用于视频信号解码的编码器 216 的方框图。

在如图 1 所示的通信系统中，一个需要发送的多媒体信号被加在第一平台 3 中的编码器 1 上。编码器 1 用于从输入信号中推出编码多媒体信号。编码器 1 的输出与发送器 2 的一个输入相连接。发送器 2 用于推出适合于发送的发送信号。发送器的输出构成了第一平台的输出，而且它与包交换传输网 4 相连接。

第二平台 6 也与包交换网 4 相连接。第二平台 6 包括用于从网络 4 接收包含有编码多媒体信号的包的接收器 8。接收器 8 将包含有编码多媒体信号的包传递给缓冲存储器 10。一般地，缓冲存储器 10 是一个 FIFO 存储器，在该存储器中可以按照与将包写入缓冲存储器 10 时的相同次序从缓冲存储器 10 中读出包。携带有暂时保存在缓冲存储器 10 中的缓冲包的缓冲存储器 10 的第一输出与演示设备 14 相连接。

30 携带有表征携带多媒体信号的包的到达延迟测量结果的缓冲存储器 10 的第二输出，与控制器件 12 的第一输入相连接。表征到达延迟的测量结果可以包括当前在缓冲器中的包的数目。如果延迟增加，当

前在缓冲存储器 10 中的包数目将减少，当延迟减小，当前缓冲器中的包数目将增加。通过计算读出指针和写入指针的位置之间的差值，便可容易地确定当前在缓冲器中的包的数目。

5 如果多媒体信号中包括时间标记，也可以通过比较与多媒体信号中预先确定的部分相关联的时间标记和多媒体信号中上述预先确定的部分的实际到达时间来推出延迟测量结果。

10 携带有读出控制信号的控制器件 12 的第一输出与缓冲存储器 10 的第二输出相连接。读出控制信号指令缓冲存储器 10 将下一个包提供给它的输出。携带有表征演示速度的信号的控制器件 12 的第二输出与演示设备 14 中的解码器 16 的控制输入相连接。根据本发明的发明理念，控制器件 12 根据表征传输延迟的测量结果确定演示速度。这里传输延迟的这个测量结果是当前缓冲存储器 10 中的包的数目。节长指标向编码器 16 提供将要合成的节的实际长度。

15 解码器 16 从接收自缓冲存储器 10 的编码信号中推出多媒体信号采样的节。节的延续时间不需要是恒定的，可以根据节长指标而变化，以便改变多媒体信息的演示速度。解码器 16 的输出与演示器件 18 相连接，在多媒体信号只包括音频信号的情况下，演示器件 18 可以是一个扬声器，而当多媒体信号包括视频信号时，演示器件 18 可以是一个显示器件。

20 在如图 2 所示的控制器件 12 中，一个表征传输延迟的输入信号被加在比较器 20 的第一输入上。在本具体实施例中，这个输入信号表征的是在缓冲器中的包的数目。比较器 20 将在缓冲器中的包的数目与参考值 REF 作比较。比较器 20 的输出通过低通滤波器 22 耦合到时钟信号发生器 24 的控制输入中。时钟信号发生器 24 产生缓冲器 10
25 的读出控制信号和解码器 16 的帧长指标。

如果在缓冲器中的包的数目小于参考值，则意味着传输延迟已经增加了。因此比较器 20 产生一个输出信号，该信号可使时钟信号发生器减小读出控制信号的频率，并增加由帧长指标标记的帧长。这将降低演示速度。由于演示速度降低，从缓冲器读出的内容便少了，因
30 而有机会将缓冲器用包充满。因此，在一定时间后缓冲器中的包的数目将增加。

如果在缓冲器中的包的数目超过了参考值 REF，比较器将产生一

个输出信号，该信号可使时钟信号发生器增加控制信号的频率，并减小由帧长指标标记的帧长。超出参考值，举例来说，可能是由传输延迟的突然减小引起的。读出控制信号频率的降低将导致演示速度的增加。由于演示速度增加，在一定时间后在缓冲器中的包的数目将减小。

在这种方法中，可以得到一个通过相应地改变演示速度来补偿延迟变化的控制循环。位于比较器 20 和时钟信号发生器之间的滤波器 22 可以在将比较器的输出信号加在时钟信号发生器上之前对其作一些平滑。去掉滤波器 22 也是可行的。

10 为了实现用缓冲器 10 中最小的延迟来补偿延迟变化，参考值 REF 可以作为（平均）延迟扩展的函数而变化。

如果由于传输通道显示几乎没有延迟扩展而使演示速度几乎是恒定的，缓冲器的尺寸可以非常小。在这种情况下，可以将参考值设置为一个低的值。

15 如果由于传输通道显示出显著的延迟扩展而使演示速度出现大的变化，缓冲器的尺寸应该大一些，以防止缓冲器变空。在这种情况下，参考值应该设置为明显高一些的值。

使参考值依赖于演示速度的变化，便可使用与延迟扩展相对应的缓冲器尺寸。这些措施可以得到低的端到端延迟，并且在多媒体信号中不会有可察觉的呃喘。

通过计算延迟测量结果的最大值和最小值之间的差值，就可以容易地确定延迟扩展。该最大和最小延迟值都是在一个给定测量时间段内确定的。

25 也可以多媒体信号回放的开始阶段将参考值设置为一个低的值，以便得到快速响应。在这种方法中，可以减少响应时间到几十个包的延续时间，其值相当于 $\pm 200\text{ms}$ 。

在如图 3 所示的控制器 12 的另一个具体实施例中，假定每个包都包含有一个时间标记。使用计数器 353 可以从由时钟谐振器 352 产生、也决定了演示速度的时钟信号中推出模拟时间标记。加法器 350 确定包的实时时间标记和可从计数器 353 输出得到的模拟时间标记之间的差值。根据本发明的发明理念，该差值就是延迟测量结果。

如果实际时间标记大于模拟时间标记，则演示速度低于新包的到

达速度。为了防止缓冲器溢出，应增加演示速度。如果实际时间标记小于模拟时间标记，则演示速度高于新包的到达速度。为了防止缓冲器变空，应降低演示速度。低通滤波器 351 用来平滑演示速度的变化。下面给出由接收速率 f_r 推出演示速率 f_p 的另一种算法。接收速率 f_r 由 $1/(T_{\text{receive}}[k]-T_{\text{receive}}[k-1])$ 确定，其中 $T_{\text{receive}}[k]-T_{\text{receive}}[k-1]$ 是两个相邻包的到达时间之差。演示速率 f_p 由 $1/(T_{\text{presentation}}[k]-T_{\text{presentation}}[k-1])$ 确定，其中 $T_{\text{presentation}}[k]-T_{\text{presentation}}[k-1]$ 是两个相邻包的演示时间之差。

以下假定，两个相邻包的到达时间差分值永不大于前两个到达时间差分值之和。这可写作：

$$10 \quad \forall i: \frac{1}{f_i[i]} < \frac{1}{f_i[i-1]} + \frac{1}{f_i[i-2]} \quad (1)$$

该算法的目的是在缓冲器中保持有 3 个包。该算法的运算如下：

A. 如果在 $T_p[i-2]$ 时刻，在缓冲器中有三个包（包 $i-2$ ，包 $i-1$ 和包 i ），包 $i-2$ 被从缓冲器中取出并按照前一个包 $i-3$ 的接收速率演示给用户。这可以表示为 $f_p[i-2]=f_r[i-3]$ 。

15 B. 在 $T_p[i-1]$ 时刻，包 $i-2$ 的演示已经完成。 $T_p[i-1]$ 可写作：

$$T_p[i-1] = t_p[i-2] + \frac{1}{f_p[i-2]} = t_p[i-2] + \frac{1}{f_r[i-3]} \quad (2)$$

现在可以区分两种情况。如果在 $T_p[i-1]$ 时刻包 $i+1$ 已经到达，缓冲器中又有了三个包，因此下一个包 $i-1$ 的演示速率由 A 确定。在包 $i+1$ 还没有到达，因而后续的 $f_r[i]$ 尚未知的情况下，约束包 $i+1$ 到达时刻 $T_R[i+1]$ 的假定 (1) 至少满足：

$$20 \quad T_R[i-1] = T_R[i] + \frac{1}{f_r[i]} \leq T_p[i-2] + \frac{1}{f_r[i]} < T_p[i-2] + \frac{1}{f_i[i-1]} + \frac{1}{f_i[i-2]} \quad (3)$$

在此情况下，包 $i-1$ 被从缓冲器中取出，并按照以下速率演示：

$$\frac{1}{f_p[i-1]} = \frac{1}{f_i[i-2]} + \left(\frac{1}{f_i[i-1]} + \frac{1}{f_i[i-3]} \right) \quad (4)$$

包 $i-1$ 的演示速率为由一个拉伸项进行延伸了的前一个包的接收速率。

C. 在 $T_p[i]$ 时刻，包 $i-1$ 的演示已经完成。 $T_p[i]$ 等于：

$$\begin{aligned}
T_p[i] &= T_p[i-1] + \frac{1}{f_p[i-1]} \\
&= \left(T_p[i-2] + \frac{1}{f_r[i-3]} \right) + \left(\frac{1}{f_r[i-2]} + \frac{1}{f_r[i-1]} + \frac{1}{f_r[i-3]} \right) \quad (5) \\
&= T_p[i-2] + \frac{1}{f_r[i-2]} + \frac{1}{f_r[i-1]}
\end{aligned}$$

包 i 依旧在缓冲器中等待。根据 (3)，在 $T_p[i]$ 时刻至少包 $i+1$ 也已到达。根据在缓冲器中是否还有两个或更多的包，下一个包的演示速率由 A（三个包或更多）或 B（两个包）确定。

如果假定 (1) 成立，该算法保证了缓冲器永远不会下溢。它不能约束缓冲器溢出。可以设想几种替代方法。

推导缓冲器中有三个包时的规律。假定平均而言，包以恒定的速率到达，缓冲器将会稳定，而 f_p 就等于 f_r 。

10 $f_p[i] = f_r[i]$ ，也就是说， $\Delta T_{BUF} = \text{常数}$ 。当接收速率下降时，缓冲器将会变空；否则它将保持恒定。

$$f_p[i] = \max\{f_p[i-1], f_r[i], f_r[i+1], \dots\}$$

其中 $f_p[i]$ 是缓冲器内所有包的所有 f_r 的平均值，它将输出速率稳定到一个恒定的比特速率。

15 在缓冲器中的包数目增加时，使用一个收缩项提高演示速率。

如图 4 所示的语音编码器 1 的输入信号 $s_i[n]$ 由 DC 陷波滤波器 210 进行滤波，以消除来自输入的不希望的 DC 残留误差。上述 DC 陷波滤波器 210 具有 15Hz 的截止频率 (-3dB)。DC 陷波滤波器 210 的输出信号被加在缓冲器 211 的输入上。根据本发明，缓冲器 211 将由 400 个已经过 DC 滤波的语音采样构成的块提供给有声语音编码器 216。
上述由 400 个采样构成的块包含 5 个由 10ms 语音构成的帧（每帧包括 80 个采样）。它包括当前要被编码的帧，两个前面的帧和两个后续帧。在每个帧间隔中，缓冲器 211 将包括 80 个采样的最新接收到的帧提供给 200Hz 高通滤波器 212 的输入。高通滤波器 212 的输出与
25 无声语音编码器 214 的一个输入和有声/无声探测器 228 相连接。高通滤波器 212 将包括 360 个采样的块提供给有声/无声探测器 228，并将包括 160 个采样（如果语音合成器 4 在 5.2Kbit/sec 模式下运行）或 240 个采样（如果语音合成器 4 在 3.2Kbit/sec 模式下运行）的块提供给无

声语音编码器 214。在下表中列出了以上给出的包括多个采样的不同的块和缓冲器 211 的输出之间的关系。

元件	5.2kbit/sec		3.2kbit/sec	
	采样数	开始位置	采样数	开始位置
高通滤波器 212	80	320	80	320
有声/无声探测器 228	360	0...40	360	0...40
有声语音编码器 216	400	0	400	0
无声语音合成编 码器 214	160	120	240	120
当前要编码的帧	80	160	80	160

有声/无声探测器 228 判断当前帧是否包括有声或无声语音，并且将判断结果表征为有声/无声标志。这个标志被传递给多路复用器 222，无声语音编码器 214 和有声语音编码器 216。根据有声/无声标志的值，激活有声语音编码器 216 或无声语音编码器 214。

在有声语音编码器 216 中，输入信号被表征为多个调和相关正弦信号。有声语音编码器的输出提供了基调值，增益值和 216 个预测参数的一种表述。基调值和增益值被分别加在多路复用器 222 的对应输入上。

在 5.2kbit/sec 模式下，每 10ms 进行一次 LPC 计算。在 3.2kbit/sec 模式下，除了发生无声语音到有声语音或相反方向的转换之外，每 20ms 进行一次 LPC 计算。如果发生这种转换，在 3.2kbit/sec 模式下，也是每 10ms 进行一次 LPC 计算。

有声语音编码器输出中的 LPC 参数被传递给多路复用器 222 的对应输入中。

在无声语音编码器 14 中，增益值和 6 个预测参数被确定用来表征无声语音信号。该增益值和 6 个预测参数被传递到多路复用器 222 的对应输入中。多路复用器 222 根据有声-无声探测器 226 的判定选择编码有声语音信号或编码无声语音信号。在多路复用器 222 的输出中，可得到编码语音信号。

在如图 5 所示的语音解码器 216 中，被编码的 LPC 码和有声/无

声标志被传递给多路分用器 92。增益值和接收到的精细基调值也被传递给多路分用器 92。

如果有声/无声标志显示的是一个有声语音帧，多路分用器 92 将把精细基调值、增益和 16 个 LPC 码传递给谐波语音合成器 94。如果有声/无声标志显示的是一个无声语音帧，多路分用器 92 将把增益和 16 个 LPC 码传递给无声语音合成器 96。谐波语音合成器 94 输出的合成有声语音信号 $\hat{s}_{v,k}[n]$ 和无声语音合成器 96 输出的合成无声语音信号 $\hat{s}_{uv,k}[n]$ 都被加在多路复用器 98 对应的输入上。

在有声模式中，多路复用器 98 将谐波语音合成器 94 的输出信号 $\hat{s}_{v,k}[n]$ 传递给重叠和加法合成块 100 的输入。在无声模式中，多路复用器 98 将无声语音合成器 96 的输出信号 $\hat{s}_{uv,k}[n]$ 传递给重叠和加法合成块 100 的输入。在重叠和加法合成块 100 中，部分重叠的有声和无声语音的节被相加。重叠和加法合成块 100 的输出信号可写为：

$$\hat{s}[n] = \begin{cases} \hat{s}_{uv,k-1}[n + N_s/2] + \hat{s}_{uv,k}[n]; v_{k-1} = 0, v_k = 0, \\ \hat{s}_{uv,k-1}[n + N_s/2] + \hat{s}_{v,k}[n]; v_{k-1} = 0, v_k = 1 \\ \hat{s}_{v,k-1}[n + N_s/2] + \hat{s}_{uv,k}[n]; v_{k-1} = 1, v_k = 0 \\ \hat{s}_{v,k-1}[n + N_s/2] + \hat{s}_{v,k}[n]; v_{k-1} = 1, v_k = 1 \end{cases} \quad (6)$$

其中 $0 < n < N_s$,

在 (6) 中， N_s 是语音帧的长度， v_{k-1} 是前一个语音帧的有声/无声标志，而 v_k 是当前帧的有声/无声标志。可以看出，长度 N_s 可根据希望的演示速度而改变。如果帧 $k-1$ 的长度等于 N_{k-1} ，则 (6) 变为：

$$\hat{s}[n] = \begin{cases} \hat{s}_{uv,k-1}[n + N_{k-1}/2] + \hat{s}_{uv,k}[n]; v_{k-1} = 0, v_k = 0, \\ \hat{s}_{uv,k-1}[n + N_{k-1}/2] + \hat{s}_{v,k}[n]; v_{k-1} = 0, v_k = 1 \\ \hat{s}_{v,k-1}[n + N_{k-1}/2] + \hat{s}_{uv,k}[n]; v_{k-1} = 1, v_k = 0 \\ \hat{s}_{v,k-1}[n + N_{k-1}/2] + \hat{s}_{v,k}[n]; v_{k-1} = 1, v_k = 1 \end{cases} \quad (7)$$

其中 $0 < n < N_s$,

重叠和加法合成块 100 的输出信号 $\hat{s}[n]$ 被加在后置滤波器 102 上。后置滤波器 102 用于通过抑制共振峰区以外的噪声来提高可觉察语音质量。

在如图 6 所示的有声语音解码器 94 中，接收自多路分用器 92 的编码基调由基调解码器 104 解码并转换为基调频率。由基调解码器 104 确定的基调频率被加在相位合成器 106 的一个输入、谐振器存储体 108

的一个输入和 LPC 谱包络采样器 110 的第一输入上。

接收自多路分用器 92 的 LPC 参数由 LPC 解码器 112 解码。LPC 参数的解码方法取决于当前的语音帧中是否包含有声或无声语音。因此有声/无声标志被加在 LPC 解码器 112 的第二输入上。LPC 解码器
5 将重构的 a -参数传递给 LPC 谱包络采样器 110。由于在精细基调计算器 32 中也要执行相同的操作，所以 LPC 谱包络采样器 112 的运行由 (13) (14) 和 (15) 描述。

相位合成器 106 用于计算表征语音信号的 L 信号的第 i 个正弦信号的相位 $\phi_k[i]$ 。 $\phi_k[i]$ 的选取要求从一个帧到下一个帧第 i 个正弦信号
10 保持连续。有声语音信号是通过将互相重叠的帧结合在一起的方法合成的，每个都包含有 N_s 个窗口采样。从图 7 中的曲线 219 和曲线 223 可以看出，两个相邻帧之间有 50% 的重叠。曲线 219 和曲线 223 中使用的窗口用点划线标示。现在相位合成器用于在重叠的相互影响最大的位置处提供一个连续相位。对于这里使用的窗口函数，这个位置位
15 于采样 119 处。当前帧的相位 $\phi_k[i]$ 可以写作：

$$\phi_k[i] = \phi_{k-1}[i] + i \cdot \omega_{0,k-1} \frac{3N_s}{4} - i \cdot \omega_{0,k-1} \frac{N_s}{4}, 1 \leq i \leq 100 \quad (8)$$

在现在描述的语音编码器中， N_s 的值等于 160。对于最初第一个有声语音帧， $\phi_k[i]$ 的值被始化为一个预先确定的值。

谐波振荡器存储体 108 产生表征语音信号的多个调和相关信号
20 $\hat{s}_{v,k}[n]$ 。该计算是使用谐波振幅 $\hat{m}[i]$ ，频率 \hat{f}_0 和合成相位 $\hat{\phi}[i]$ 根据下式执行的：

$$\hat{s}_{v,k}[n] = \sum_{i=1}^L \hat{m}[i] \cos\{i \cdot 2\pi \cdot \hat{f}_0 \cdot n + \hat{\phi}[i]\}; 0 < n < N_s \quad (9)$$

在时域窗口化块 114 中使用 Hanning 窗口将信号 $\hat{s}'_{v,k}[n]$ 窗口化。该窗口信号显示在图 7 的曲线 221 中。使用具有在时间上有 $N_s/2$ 采样移
25 动的 Hanning 窗口将信号 $\hat{s}'_{v,k+1}[n]$ 窗口化。该窗口信号显示在图 7 的曲线 225 中。将上述窗口信号相加，便得到时域窗口化块 114 的输出信号。该输出信号显示在图 7 的曲线 227 中。增益解码器 118 从其输入信号推出增益值 g_v ，而时域窗口化块 114 的输出信号由信号标定块 116 用上述增益因子 g_v 标定，以便得到重构的有声语音信号 $\hat{s}_{v,k}[n]$ 。

根据本发明的发明理念, 如果改变了多媒体的演示速度, 则应对上述合成过程作一些变化。以下假定帧长指标由数个采样 N_i 表征, 其中 i 是帧的数目。首先, 需要由在要合成的当前帧之前的帧的采样数目 N_{i-1} 和 N_{i-2} 确定相位 $\varphi_k[i]$ 。根据下式计算这些相位:

$$5 \quad \varphi_k[i] = \varphi_{k-1}[i] + i \cdot 2\pi \cdot f_{0,k-1} \left[\frac{N_{k-2}}{2} + \frac{N_{k-1}}{4} \right] - i \cdot 2\pi \cdot f_{0,k} \frac{N_{k-1}}{4}; 1 \leq i \leq 100 \quad (10)$$

随后根据下式合成信号 $\hat{s}'_{v,k+1}[n]$:

$$\hat{s}'_{v,k}[n] = \sum_{i=1}^L \hat{m}[i] \cos\{i \cdot 2\pi \cdot f_0\} \cdot n + \hat{\phi}[i]; 0 < n < N_s \quad (11)$$

当帧中的采样数目与标称值 N_s 不同时, 时域窗口化块 114 的操作也稍有改变。用于将信号 $\hat{s}'_{v,k}[n]$ 窗口化的 Hanning 窗口的长度等于 N_k , 而不是 N_s 。

在图 8 中显示有与图 7 相同的信号, 但是现在两个节的边界处的演示速度改变了。曲线 418 表征的节明显短于曲线 422 表征的节。将曲线 420 和 424 所示的窗口信号窗口化并相加后, 得到曲线 426 所示的信号。

15 在如图 9 所示的无声语音合成器 96 中, LPC 码和有声/无声标志被加在 LPC 解码器 130 上。LPC 解码器 130 向 LPC 合成滤波器 134 提供多组 6 个 a -参数。高斯白噪声发生器 132 的输出与 LPC 合成滤波器 134 的一个输入相连接。LPC 合成滤波器 134 的输出信号在时域窗口化块 140 中由 Hanning 窗口窗口化。

20 无声增益解码器 136 推出表征当前无声帧的希望的能量的增益值 \hat{g}_{uv} 。从该增益和窗口信号的能量可确定窗口化语音信号增益的标定因子 \hat{g}'_{uv} , 以得到具有正确能量的语音信号。该标定因子可写为:

$$\hat{g}'_{uv} = \sqrt{\frac{\hat{g}_{uv}}{\frac{1}{N_s} \sum_{n=0}^{N_s-1} (\hat{s}'_{uv,k}[n] \cdot w[n])^2}} \quad (12)$$

25 信号标定块 142 通过将时域窗口块 140 的输出信号乘以标定因子 \hat{g}'_{uv} 来确定输出信号 $\hat{s}_{v,k}[n]$ 。

可改动当前描述的语音编码系统, 以使之需要较低的比特率或较高的语音质量。需要较低比特率的语音编码系统的一个例子是 2kbit/sec 编码系统。将用于有声语音的预测参数的数目从 16 减少到

12, 并使用预测参数, 增益和精细基调的差分编码方法, 便可得到这样一个系统。差分编码意味着不单独对需要编码的数据进行编码, 而是只发送连续帧对应数据之间的差值。在从有声到无声的语音转换或相反方向的转换中, 在第一个新的帧中, 所有参数都被单独编码以便
5 为解码提供初始值。

也可以得到能够在 6kbit/s 比特率下提高语音质量的语音编码器。这里的改动是多个调和相关正弦信号的前 8 个谐波的相位确定。相位 $\varphi[i]$ 按照下式计算:

$$\varphi[i] = \arctan \frac{I(\theta_i)}{R(\theta_i)} \quad (13)$$

10 这里 $\theta_i = 2\pi f_0 \cdot i$, $R(\theta_i)$ 和 $I(\theta_i)$ 等于

$$R(\theta_i) = \sum_{n=0}^{N-1} s_w[n] \cdot \cos(\theta_i \cdot n) \quad (14)$$

和

$$I(\theta_i) = -\sum_{n=0}^{N-1} s_w[n] \cdot \sin(\theta_i \cdot n) \quad (15)$$

如此得到的 8 个相位值 $\varphi[i]$ 统一量化到 6 比特, 并包括在输出比特
15 流中。

6kbit/sec 编码器的更进一步的改动是无声模式中额外增益值的传输。通常每 2ms 传输一个增益值, 而不是每帧传输一次。在转换后的第一帧中, 要传输 10 个增益值, 其中 5 个表征当前无声帧, 而另 5 个表征由无声语音编码器处理的前一个有声帧。增益是从 4ms 的重叠
20 窗口中确定的。

在如图 10 所示的视频编码器 16 中, 携带有由多个视频帧组成的视频信号耦合到插入器 304 的第一输入和帧存储器 302 的输入中。帧存储器 302 用于存储先前接收自缓冲器 10 的视频帧。帧存储器 302 的输出与插入器 304 的第二输入相连接。

25 插入器 304 用于内插接收自缓冲器 10 的前一个视频帧和当前的视频帧。插入器向自己的输出提供具有恒定帧速率的视频信号, 供演示器件 18 使用。

根据本发明的发明理念, 演示速度取决于延迟测量结果。在此情况下, 这意味着接收自缓冲器 10 的视频帧并不一直以相等间隔显示。

两帧之间的间隔取决于延迟测量结果。

为了能够以大体恒定的帧速率演示视频信号，插入器 304 确定了取决于接收自缓冲器 10 的视频帧之间间隔的数个内插帧。

5 计算设备 306 根据由图 2 中的时钟发生器 24 提供的演示速度计算出需要内插的帧的数目。在视频信号中使用了时间标记的情况下，当前帧和前一个帧的时间标记之间的差值 Δ 被提供给计算设备 306。这使计算设备 306 在一个或多个视频帧丢失时也可以确定需要内插的帧的正确数目。

10 在 1998 年 3 月在奥兰多举行的 Winhec98 大会上，G. De Haan 在文章“个人电脑中的无抖动视频”中描述了一款适用的内插器 304。

说明书附图

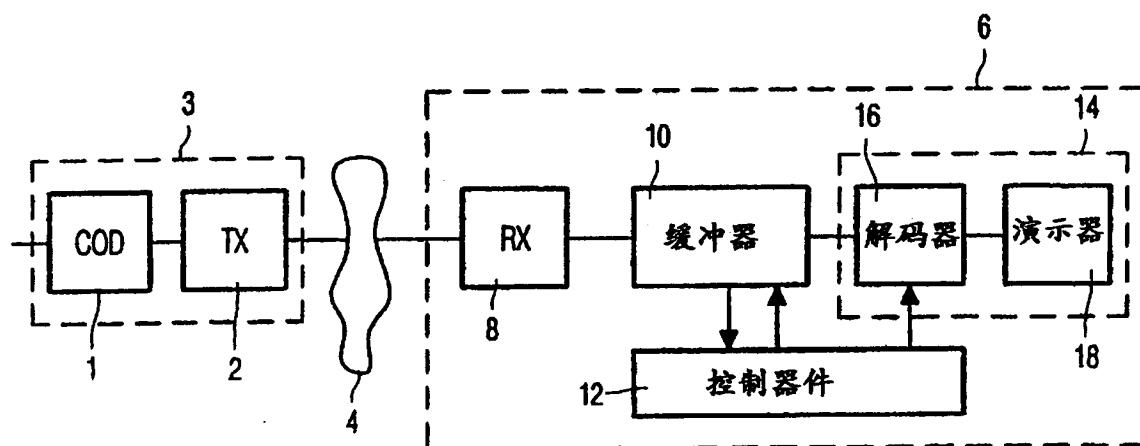


图 1

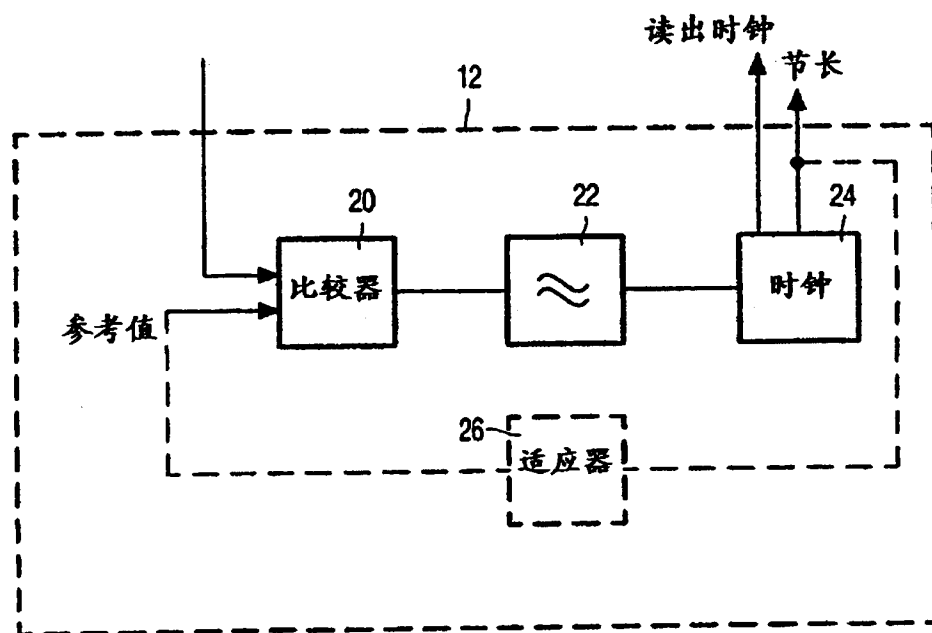


图 2

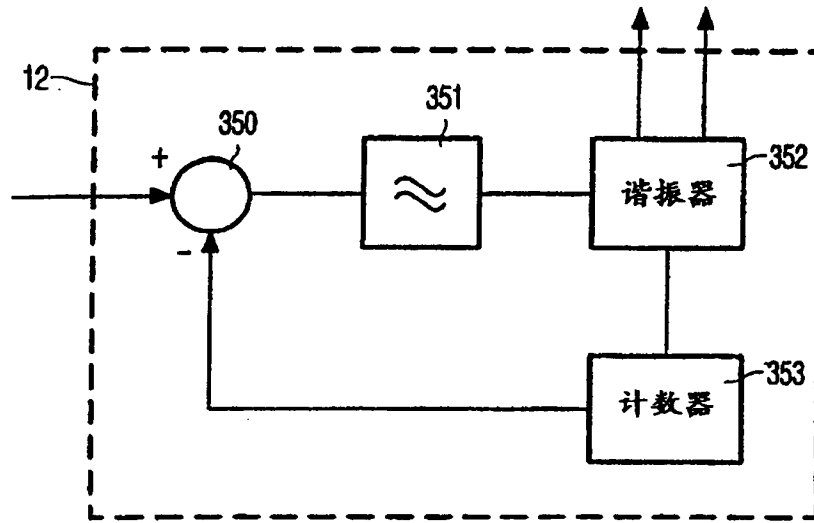


图 3

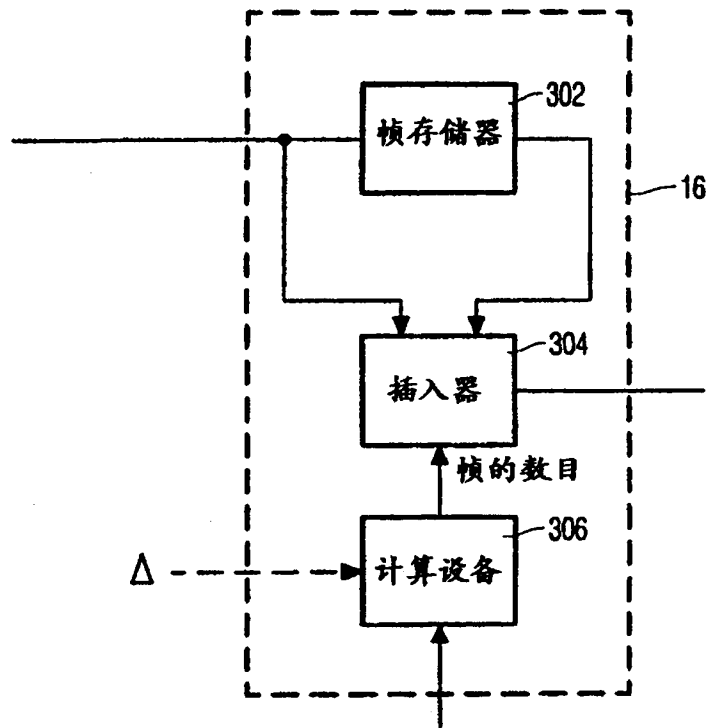


图 10

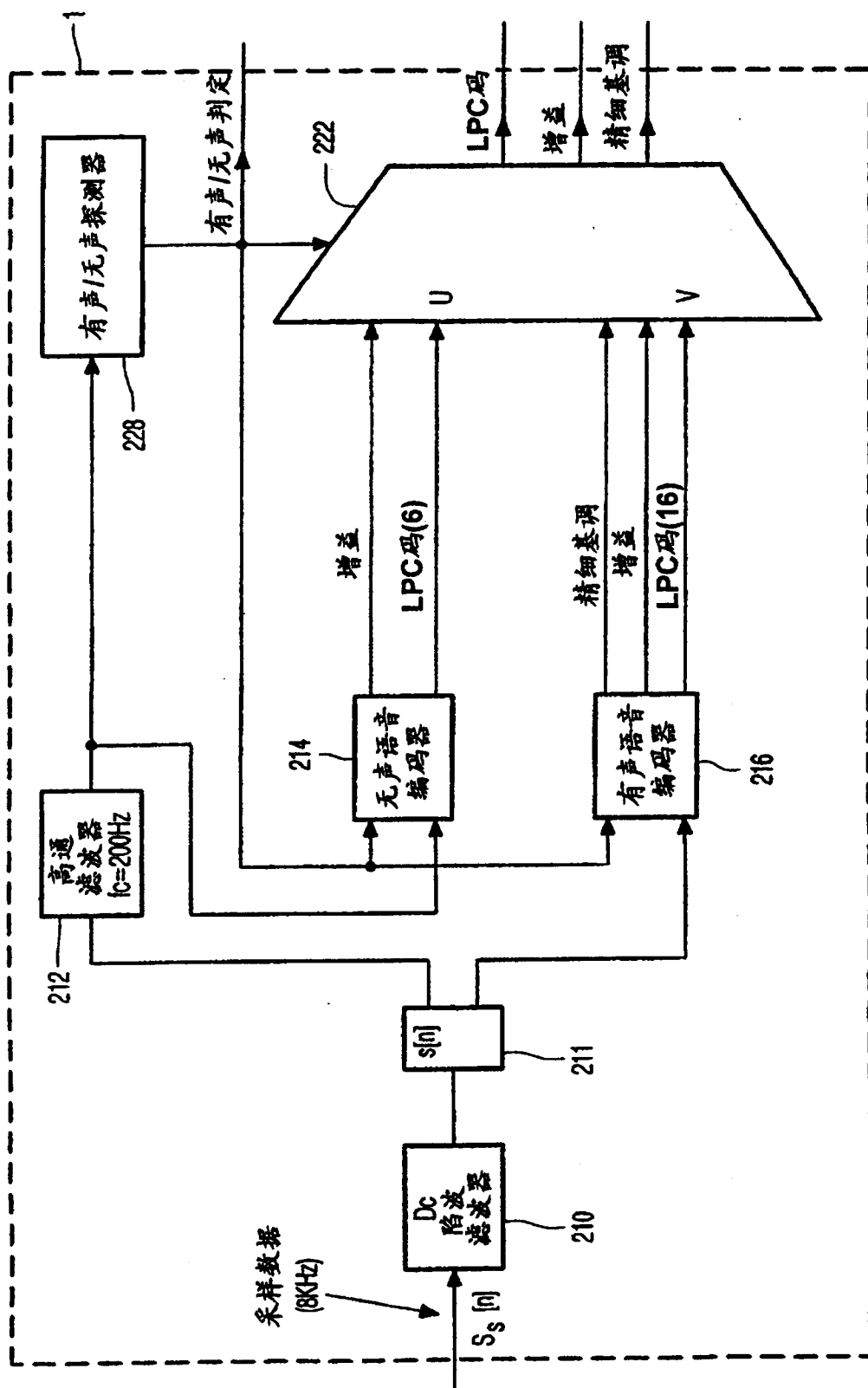


图 4

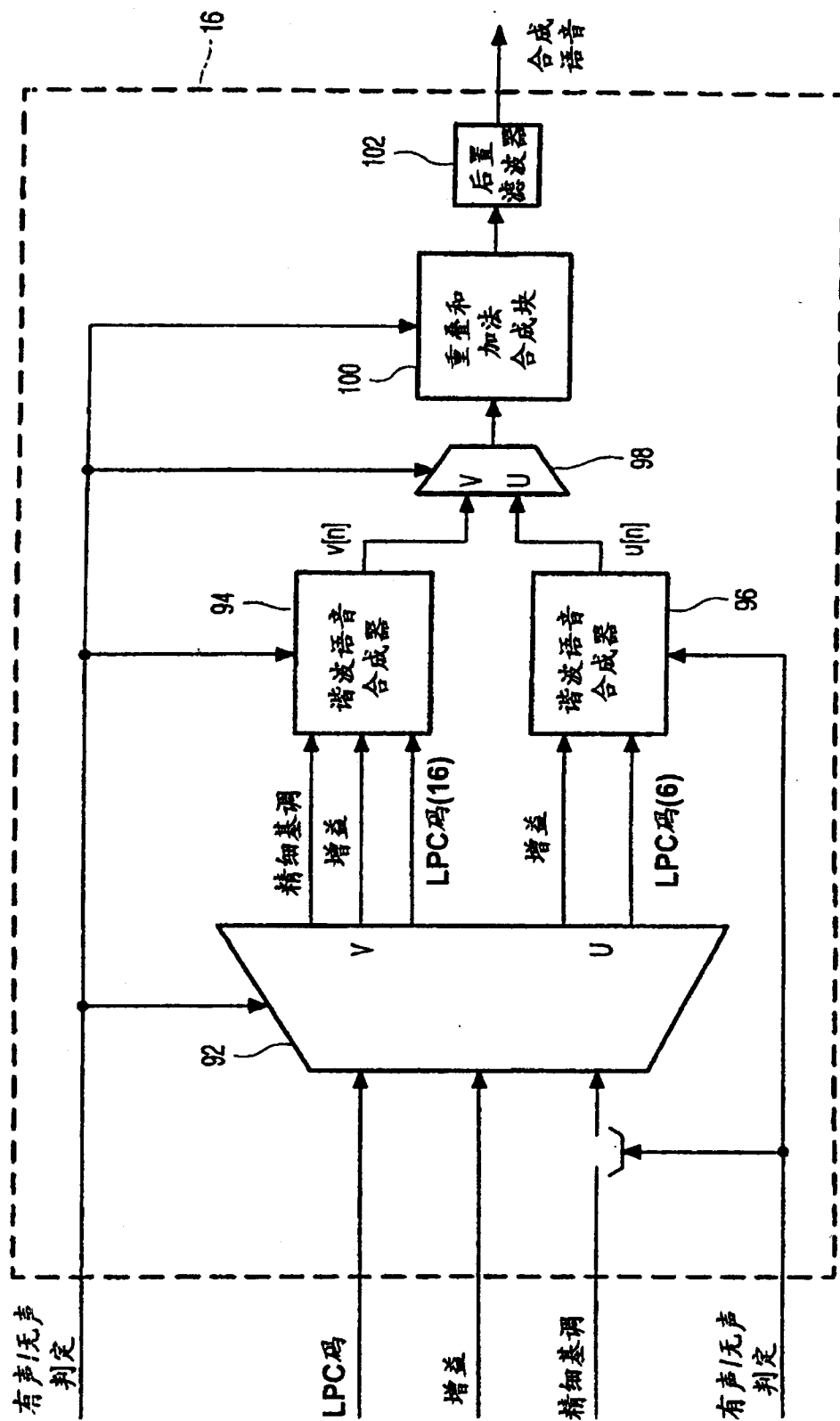


图 5

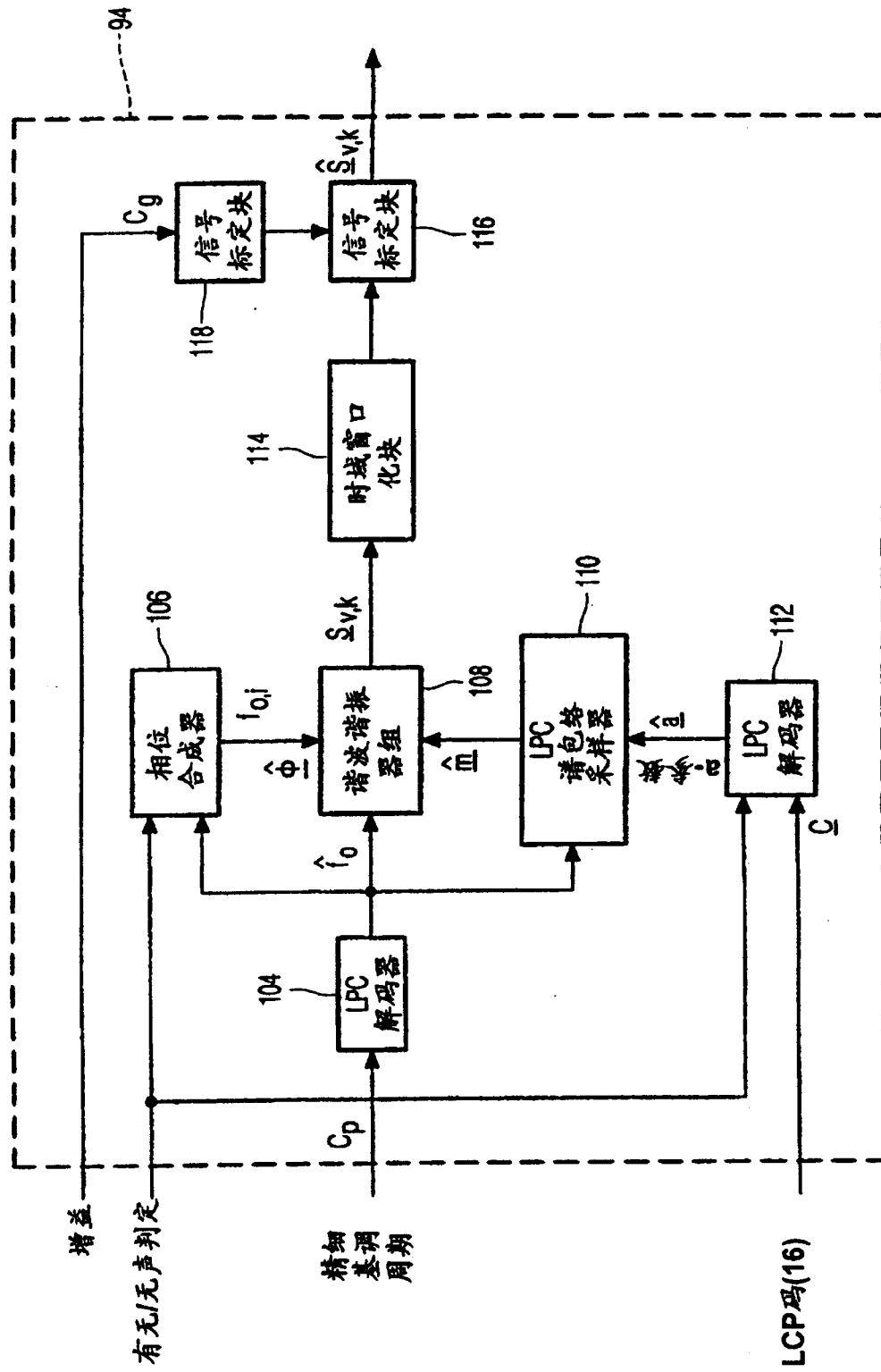


图 6

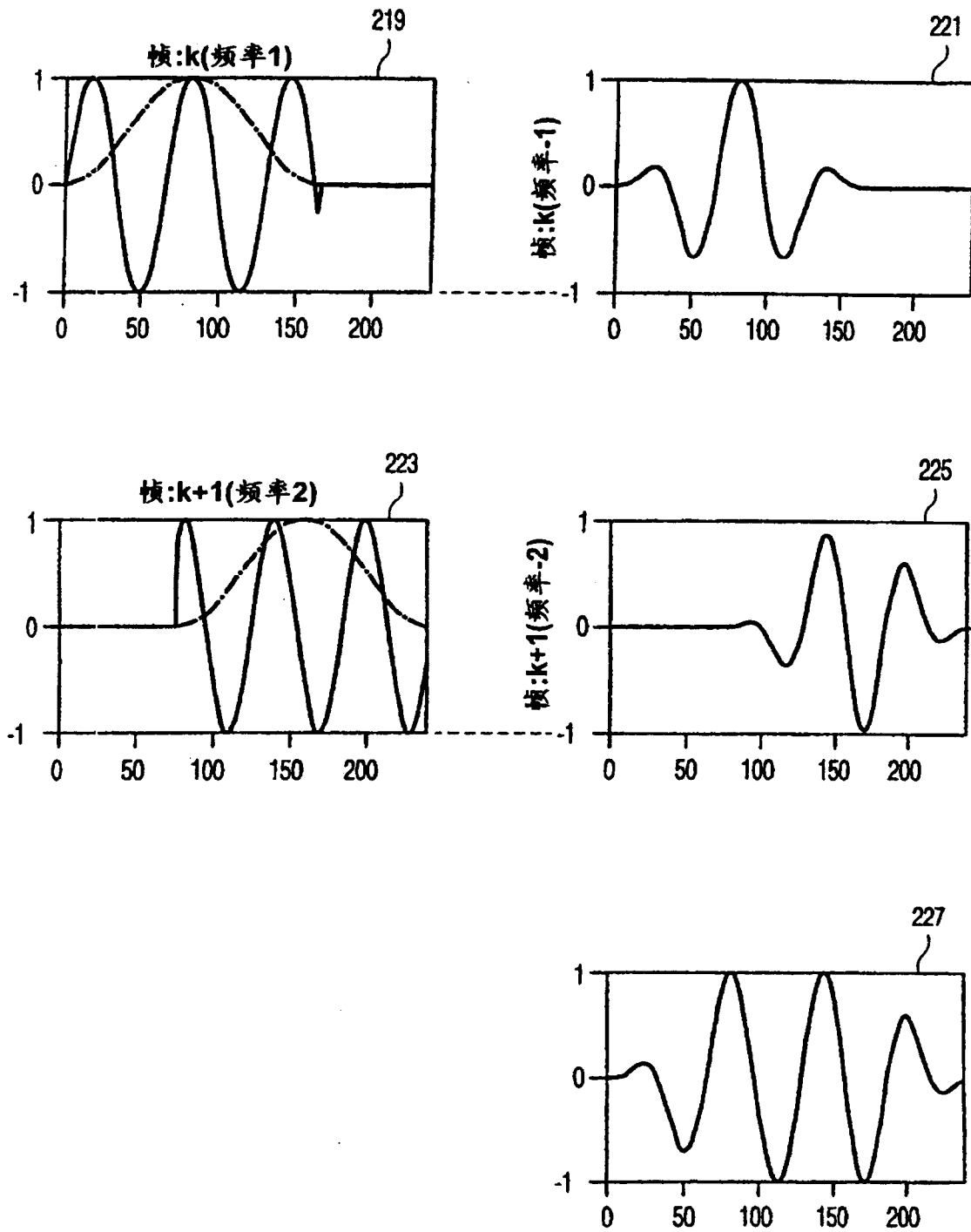


图 7

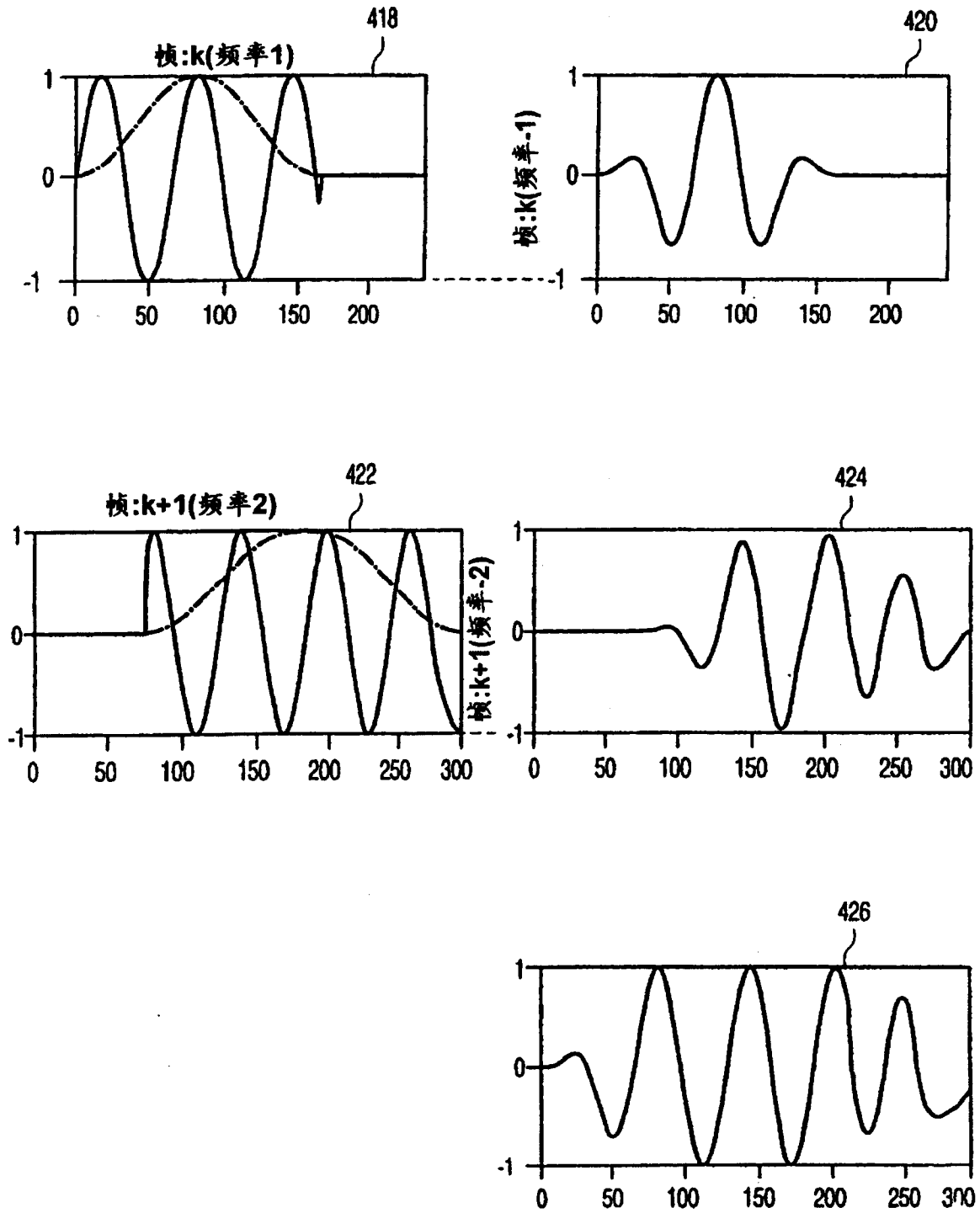


图 8

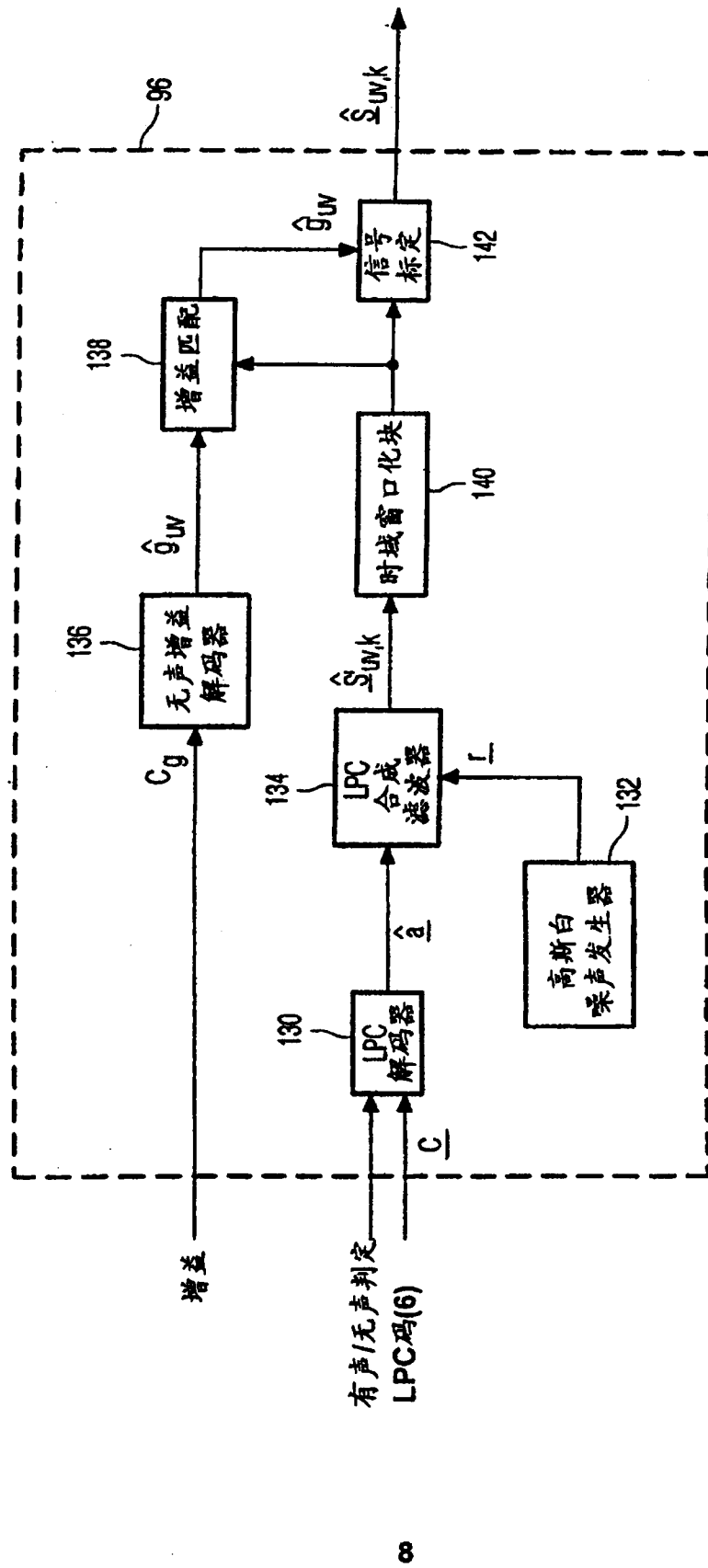


图 9